

10-6-2021

The Aquatic Symbiosis Genomics Project: probing the evolution of symbiosis across the tree of life

Victoria McKenna

John M. Archibald

Roxanne Beinart

Michael N. Dawson

Ute Hentschel

See next page for additional authors

Authors

Victoria McKenna, John M. Archibald, Roxanne Beinart, Michael N. Dawson, Ute Hentschel, Patrick J. Keeling, Jose V. Lopez, José M. Martín-Durán, Jillian M. Petersen, Julia D. Sigwart, Oleg Simakov, Kelly R. Sutherland, Michael Sweet, Nick Talbot, Anne W. Thompson, Sara Bender, Peter W. Harrison, Jeena Rajan, Guy Cochrane, Matthew Berriman, Mara Lawniczak, and Mark Blaxter



OPEN LETTER

The Aquatic Symbiosis Genomics Project: probing the evolution of symbiosis across the tree of life [version 1; peer review: awaiting peer review]

Victoria McKenna¹, John M. Archibald², Roxanne Beinart³, Michael N. Dawson⁴, Ute Hentschel⁵, Patrick J. Keeling⁶, Jose V. Lopez ⁷, José M. Martín-Durán⁸, Jillian M. Petersen⁹, Julia D. Sigwart ¹⁰, Oleg Simakov¹¹, Kelly R. Sutherland ¹², Michael Sweet¹³, Nick Talbot¹⁴, Anne W. Thompson ¹⁵, Sara Bender¹⁶, Peter W. Harrison ¹⁷, Jeena Rajan¹⁷, Guy Cochrane¹⁷, Matthew Berriman ¹, Mara Lawniczak¹, Mark Blaxter ¹

¹Tree of Life, Wellcome Sanger Institute, Cambridge, CB10 1SA, UK

²Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia, B3H 4R2, Canada

³Graduate School of Oceanography, University of Rhode Island, Narragansett, Rhode Island, 02882, USA

⁴Life & Environmental Sciences, University of California, Merced, Merced, California, 95343, USA

⁵Marine Symbioses Research Unit, GEOMAR Helmholtz Centre for Ocean Research, Kiel, Germany

⁶Department of Botany, University of British Columbia, Vancouver, British Columbia, V6T 1Z4, Canada

⁷Halmos College of Arts and Sciences, Nova Southeastern University, Dania Beach, Florida, 33004, USA

⁸School of Biological and Chemical Sciences, Queen Mary University of London, London, E1 4NA, UK

⁹Centre for Microbiology and Environmental Systems Science, University of Vienna, Vienna, A-1090, Austria

¹⁰Marine Zoology Department, Senckenberg Research Institute, Frankfurt, 60325, Germany

¹¹Department for Neurosciences and Developmental Biology, University of Vienna, Vienna, 1010, Austria

¹²Department of Biology, University of Oregon, Eugene, Oregon, 97403-1210, USA

¹³Aquatic Research Facility, Environmental Sustainability Research Centre, University of Derby, Derby, DE22 1GB, UK

¹⁴The Sainsbury Laboratory, Norwich, NR4 7UH, UK

¹⁵Department of Biology, Portland State University, Portland, Oregon, 97201, USA

¹⁶Gordon and Betty Moore Foundation, Palo Alto, California, 94304, USA

¹⁷European Molecular Biology Laboratory, European Bioinformatics Institute, Cambridge, CB10 1SD, UK

V1 First published: 06 Oct 2021, 6:254
<https://doi.org/10.12688/wellcomeopenres.17222.1>

Latest published: 06 Oct 2021, 6:254
<https://doi.org/10.12688/wellcomeopenres.17222.1>

Abstract

We present the Aquatic Symbiosis Genomics Project, a global collaboration to generate high quality genome sequences for a wide range of eukaryotes and their microbial symbionts. Launched under the Symbiosis in Aquatic Systems Initiative of the Gordon and Betty Moore Foundation, the ASG Project brings together researchers from across the globe who hope to use these reference genomes to augment and extend their analyses of the dynamics, mechanisms and environmental importance of symbiosis. Applying large-scale, high-throughput sequencing and assembly technologies, the ASG

Open Peer Review

Reviewer Status AWAITING PEER REVIEW

Any reports and responses or comments on the article can be found at the end of the article.

collaboration will assemble and annotate the genomes of 500 symbiotic organisms – both the “hosts” and the microbial symbionts with which they associate. These data will be released openly to benefit all who work on symbiosis, from conservation geneticists to those interested in the origin of the eukaryotic cell.

Keywords

Symbiosis, Marine, Freshwater, Genome Sequencing, Collaboration, Open Science

Corresponding author: Mark Blaxter (mark.blaxter@sanger.ac.uk)

Author roles: **McKenna V:** Conceptualization, Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing; **Archibald JM:** Conceptualization, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Beinart R:** Conceptualization, Writing – Review & Editing; **Dawson MN:** Conceptualization, Writing – Review & Editing; **Hentschel U:** Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Keeling PJ:** Conceptualization, Writing – Review & Editing; **Lopez JV:** Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Martín-Durán JM:** Conceptualization, Writing – Review & Editing; **Petersen JM:** Conceptualization, Writing – Review & Editing; **Sigwart JD:** Conceptualization, Writing – Review & Editing; **Simakov O:** Conceptualization, Writing – Review & Editing; **Sutherland KR:** Conceptualization, Writing – Review & Editing; **Sweet M:** Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Talbot N:** Conceptualization, Writing – Review & Editing; **Thompson AW:** Conceptualization, Writing – Review & Editing; **Bender S:** Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Harrison PW:** Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Rajan J:** Conceptualization, Writing – Review & Editing; **Cochrane G:** Conceptualization, Funding Acquisition, Writing – Original Draft Preparation, Writing – Review & Editing; **Berriman M:** Conceptualization, Funding Acquisition, Writing – Original Draft Preparation, Writing – Review & Editing; **Lawniczak M:** Conceptualization, Funding Acquisition, Writing – Original Draft Preparation, Writing – Review & Editing; **Blaxter M:** Conceptualization, Funding Acquisition, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by Wellcome [206194] and the Gordon and Betty Moore Foundation [GBMF8897]. *The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

Copyright: © 2021 McKenna V *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: McKenna V, Archibald JM, Beinart R *et al.* **The Aquatic Symbiosis Genomics Project: probing the evolution of symbiosis across the tree of life [version 1; peer review: awaiting peer review]** Wellcome Open Research 2021, 6:254 <https://doi.org/10.12688/wellcomeopenres.17222.1>

First published: 06 Oct 2021, 6:254 <https://doi.org/10.12688/wellcomeopenres.17222.1>

Disclaimer

The views expressed in this article are those of the author(s). Publication in Wellcome Open Research does not imply endorsement by Wellcome.

The genomics of symbiosis

Symbiosis, the living together of distinct organisms (Archibald, 2014; Oulhen *et al.*, 2016), describes a spectrum of relationships from mutualistic to parasitic, and from obligate to temporary. Symbiosis has been and is fundamental to the evolution of life on Earth, from the deep origins of the eukaryotic cell and photosynthetic eukaryotes, through to the recent emergence of new partnerships. The power of symbiosis arises from the ability of the joint organism to draw from the independent, billion-year evolutionary histories of both partners. Symbiosis is a fact of life – it has arisen many, many times

and new symbioses are constantly evolving (Figure 1). In this era of rapid climate change and biodiversity loss, many key-stone symbiotic systems are threatened, and their loss imperils the ecosystems they support.

Well-known mutualist symbioses permit colonisation of otherwise inaccessible habitats, are critical to ecosystem functioning, and support marine and freshwater diversity. For example, coral reefs, built through a photosymbiotic association between cnidarians and dinoflagellate algae (Weis, 2019), create biodiversity hotspots which house upwards of 25% of all described species in the oceans. The dominant animals colonising deep-sea hydrothermal vents are nutritionally dependent on chemosymbiotic associations with bacteria (Roeselers & Newton, 2012), allowing them to thrive in the food-limited dark ocean. For these symbioses, the biological fitness consequences

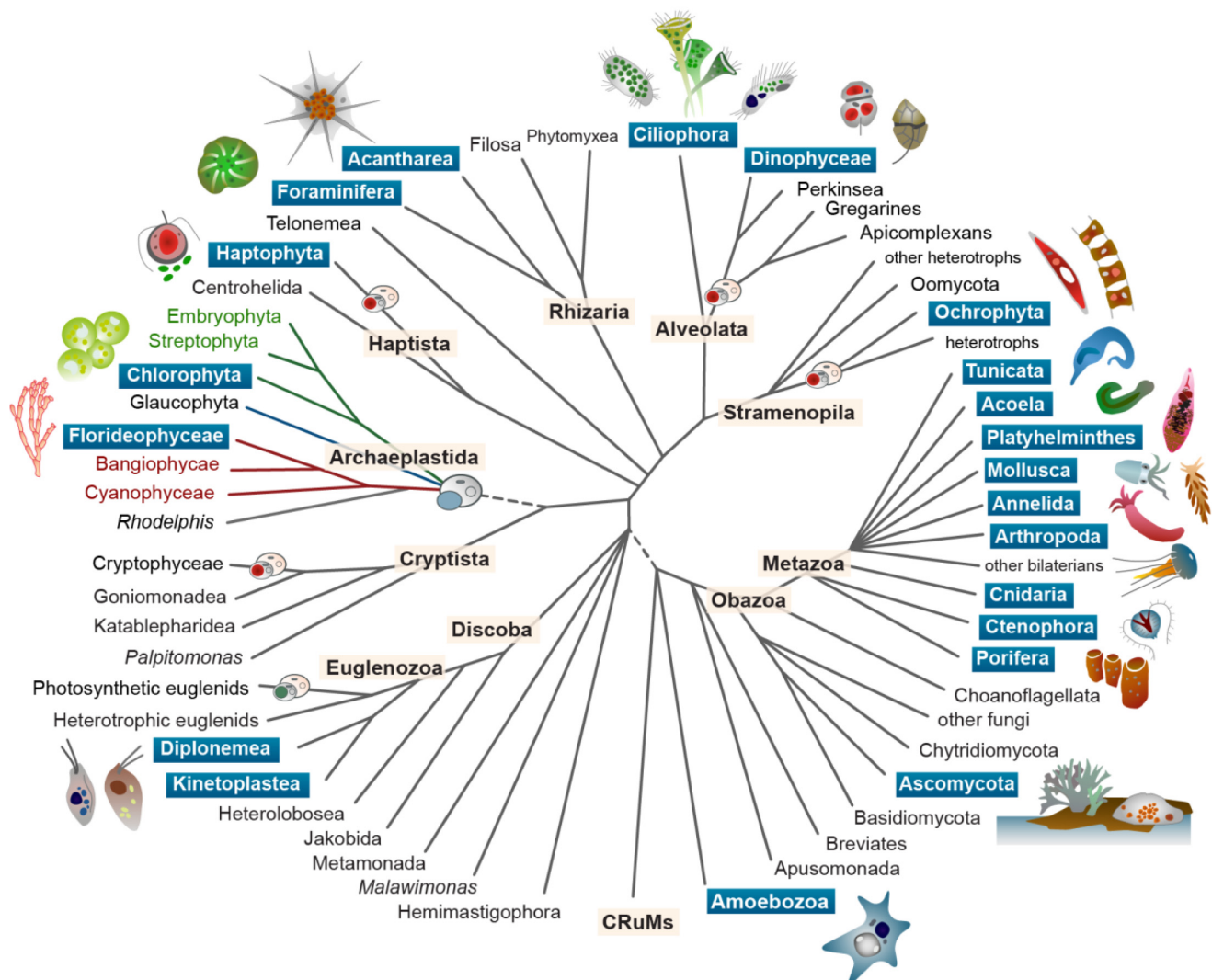


Figure 1. The phylogenetic diversity of eukaryotic symbioses. Symbiotic taxa, and Aquatic Symbiosis Genomics target species, are found across the diversity of the eukaryotic tree of life. Taxa highlighted with blue boxes include ASG targets. Within the tree, the small cartoons indicate the major event of plastid acquisition through symbiosis with a cyanobacterium (in the Archaeplastida; blue cell engulfed) and the several events of secondary and tertiary plastid acquisition in other lineages. Illustration by John Archibald and Mark Blaxter.

are largely understood, but in many less well-known symbioses, such as those between sponges and their bacterial collaborators, or partnerships in the diverse world of single celled eukaryotes, the basis of the relationships are not known in any detail.

The aquatic symbiosis genomics project will transform symbiosis research

The Gordon and Betty Moore Foundation has created a major funding initiative focused on investigating the biology of symbiosis in marine and freshwater ecosystems (see [Symbiosis in Aquatic Systems Initiative](#)). To support this global initiative, the Aquatic Symbiosis Genomics project (ASG; see [Aquatic Symbiosis Genomics Project – Wellcome Sanger Institute](#)) plans to generate high-quality genome sequences from a wide range of symbiotic systems. Our focus is on symbioses involving at least one microbial partner, and where there is likely to be co-evolving interplay between the species involved.

Like a symbiotic organism, the ASG project is more than the simple sum of its parts. ASG will merge the decades of ecological, evolutionary, taxonomic, and experimental expertise of researchers from diverse backgrounds with the decades of genomics experience of the Wellcome Sanger Institute. ASG works on a hub and spokes model, where communities of researchers nucleated on specific questions and/or species systems have come together as hubs to propose sets of taxa for

sequencing ([Table 1](#)). These (currently) total ~450 distinct symbiotic organisms from the open ocean, the deep sea, coastal, littoral, and freshwater ecosystems, which are expected to include over 1000 nominal species of hosts and symbionts. The ASG target list includes species representing many phyla of animals, protists, algae and fungi, and encompasses ancient and recently-evolved partnerships.

The hub partners have defined the major scientific questions they wish to explore, and will source and identify specimens that will deliver answers. ASG follows an ethical code of sampling practice, avoiding overcollection and respecting local and international laws and protocols, especially as ASG will be sampling from endangered ecosystems and in some cases endangered species. The project participants are fully committed to the Convention on Biological Diversity Nagoya Protocols on Access and Benefit Sharing, and only samples where express permission has been obtained will be sourced and sequenced. Samples may come from the wild, from mesocosms and aquaria, from explant lab cultures or from culture collections.

Genome sequencing and assembly will be delivered by the Tree of Life programme at the [Sanger Institute](#) using pipelines being developed for the [Darwin Tree of Life](#) and other major biodiversity genomics projects. Genomes will be assembled, annotated and released openly through the European Bioinformatics Institute ([EMBL-EBI](#)).

Table 1. Aquatic symbiosis genomics project hubs.

Lead researcher*	Project Title (short)	Major taxa represented	
		Hosts	Symbionts
Archibald	New symbioses in single-celled eukaryotes	Amoebozoa, Dinophyceae, Diplonemea (Euglenozoa), Haptophyta, Ochrophyta	Bacteria, Kinetoplastea, Ochrophyta
Beinart, Petersen, Sigwart	Molluscan symbioses	Mollusca	Arthropoda, Bacteria, Chlorophyta, Cnidaria, Dinophyceae, Platyhelminthes, Florideophyceae
Dawson, Sutherland, Thompson	Pelagic symbioses	Acoela, Ctenophora, Cnidaria, Tunicata	Bacteria, Chlorophyta, Dinophyceae, other Alveolata
Hentschel	Sponge symbioses	Porifera	Bacteria, Archaea, Viruses, <i>Symbiodinium</i> (Dinophyceae) and others
Keeling	Symbiosis in ciliates	Ciliophora	Archaea, Bacteria, Chlorophyta, Ciliophora, Dinophyceae
Lopez	Metazoan photosymbioses	Acoela, Cnidaria, Mollusca, Porifera, Tunicata	Bacteria, Chlorophyta, Cnidaria, Dinophyceae, Haptophyta, Myxozoa
Martín-Durán	Annelid chemosymbioses	Annelida	Bacteria, Archaea
Simakov	Cephalopod symbioses	Mollusca	Bacteria, Archaea
Sweet	Coral symbioses	Cnidaria	<i>Symbiodinium</i> (Dinophyceae)
Talbot	Marine lichens	Fungi	Bacteria, Chlorophyta, ascomycete Fungi, Ochrophyta

* see author list for affiliations.

Sequencing symbionts: from sample to openly accessible genome assembly

Each ASG Hub ([Table 1](#)) has defined a set of taxa that it will sample for sequencing. We will sequence from single eukaryotic host specimens or clonal cultures rather than bulk samples whenever possible. While this can limit the mass of DNA and RNA available for sequencing, it has the very strong benefit of reducing allelic sequence complexity and enabling assembly. Importantly, we do not require that the symbiotic partners are separated before sequencing, as we will separate the host and symbiont genomes bioinformatically during assembly ([Challis *et al.*, 2020](#)).

Each sample is formally identified and associated with rich metadata describing its collection location and other environmental features. We collate and validate these metadata through the [COCO biodiversity data brokering system](#). Samples are shipped to the Sanger Institute for long DNA and RNA extraction and sequencing, with particular focus on low-input methods. We are generating a combination of long read and long range genomic data. For long reads we primarily use the Pacific Biosciences Sequel IIe circular consensus sequencing approach to generate high fidelity (HiFi) reads in the 15 to 20 kilobase range, and include Oxford Nanopore Technologies long reads where needed. For long range data we use chromatin conformation capture sequencing (known as Hi-C). These long range data generate important information that link sequences within chromosomes and organelles in the multi-kilobase to megabase range and will allow us to disentangle genomes from different species. The joint transcriptome of the symbioses will be sampled using RNA-Seq, both on Illumina short read and Pacific Biosciences long read platforms.

We have strong expectations as to what we should find in the sequence data, and what we should be assembling, but biology is full of exceptions and surprises and organisms taken from the wild are frequently found in association with other cobionts. Each symbiosis contains a community of genomes that can be viewed as a low complexity metagenome: the “host” genome and the genomes of its organelles (mitochondrion and in some cases plastid), the symbiont genome (which if it is eukaryotic contains one or more organellar genomes) and the genomes of other commensals and cobionts. We separate data into presumed organismal and organellar subsets and assemble each independently. First we identify taxonomically informative marker loci, such as small subunit ribosomal RNAs (organellar 12S, prokaryotic 16S and eukaryotic 18S), cytochrome oxidase I genes, and ribulose-1,5-bisphosphate carboxylase-oxygenase genes, in the HiFi reads and primary assembly. These tell us which taxa are likely to be present and thus which genomes we should expect to assemble. To separate the data we use intrinsic features (GC and tetranucleotide composition, read coverage, coding capacity), sequence similarity to known genomes, and Hi-C linkage information. Binning contigs and their constituent reads into distinct subsets facilitates complete assembly of each organismal and organellar genome ([Challis *et al.*, 2020](#); [Kumar & Blaxter, 2011](#)). We aim to automate this cobiont identification and binning process, as it will be of utility in analyses of all tree of life genomes: many specimens

harbour parasitic and other cobionts. Given 25- to 30-fold genome coverage in HiFi reads for each symbiont partner, we expect to generate primary assemblies with contig N50s in the multi-megabase range. The Hi-C data are used to scaffold these contigs into near-chromosomal pseudomolecules.

For each symbiotic system we will then curate the assemblies to improve accuracy ([Howe *et al.*, 2021](#)) with particular attention to correct scaffolding of nuclear chromosomes and circularisation of organellar and prokaryotic genomes, and identification of remaining complex and unresolvable repetitive regions (such as ribosomal RNA and centromeric repeats). We aim to achieve or exceed the latest Earth BioGenome Project ([Lewin *et al.*, 2018](#)) assembly standards. Curated assemblies and all raw data will be submitted to the European Nucleotide Archive (ENA) ([Harrison *et al.*, 2021](#)) and from there to the rest of the International Nucleotide Sequence Database Consortium for immediate open release. The genomes will be annotated using the RNA-Seq transcriptomic data binned by species, and the annotations released openly. We have developed an [ASG-specific data portal](#) that collates all of the data generated by the project and promotes analysis. The Aquatic Symbiosis Genomics project relies on engagement and support from the whole of the Tree of Life production genomics team and of many colleagues who are participants in the ten Hubs. Each symbiotic system will be the subject of an open access publication, a Genome Note, that credits the full team that generated the assemblies, from collectors to annotators ([Threlfall & Blaxter, 2021](#)).

Building an aquatic symbiosis genomics community

The ASG project aims to generate a lasting resource in terms of the ~1000 genomes involved in ~500 symbiotic systems. To ensure this resource results in a flourishing ecosystem of postgenomic research, we are building community and expertise through a parallel programme of training and mentoring in genomics and bioinformatics. In collaboration with [Wellcome Connecting Science and The Carpentries](#), the ASG project will deliver intensive and extensive collaborative training and investigative informatic analysis of symbiont genomes, to build collective genomics and bioinformatics capacity in the symbiosis community. Training will include core informatics, coding, and reproducible science, as well as deeper analytical dives into co-evolving genomes, detailed genome annotation, and prediction of the metabolic underpinnings of symbiotic cooperation.

Just as reefs built by corals and their symbiotic algae allow an exuberant and diverse ecology to thrive, the ASG project will build a lasting genomic foundation for flourishing and diverse analyses of symbiosis. Many of the fish that throng around coral reefs are open spawners, and their larvae spend their first weeks in the open ocean. They are recruited back to the reef because they can hear and smell it: the chatter generated by a healthy reef attracts, recruits, and builds the reef community ([Gordon *et al.*, 2019](#)). Much like a healthy reef, our hope is that the high quality genomes we produce will generate the chatter that attracts new researchers and provides a

foundation for growth of fundamental research on the nature of symbiosis and conservation of habitats where symbioses abound.

Data availability

No data are associated with this article. ASG data will be released openly in the European Nucleotide Archive.

Acknowledgements

We thank Jonathan Threlfall for assistance with manuscript editing. This research was funded in part by Wellcome [grant 206194]. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

References

- Archibald J: **One Plus One Equals One: Symbiosis and the Evolution of Complex Life**. Oxford University Press, USA. 2014.
[Reference Source](#)
- Challis R, Richards E, Rajan J, *et al.*: **BlobToolKit - Interactive Quality Assessment of Genome Assemblies**. *G3*. 2020; **10**(4): 1361–74.
[Publisher Full Text](#)
- Gordon TAC, Radford AN, Davidson IK, *et al.*: **Acoustic Enrichment Can Enhance Fish Community Development on Degraded Coral Reef Habitat**. *Nat Commun*. 2019; **10**(1): 5414.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Harrison PW, Ahamed A, Aslam R, *et al.*: **The European Nucleotide Archive in 2020**. *Nucleic Acids Res*. 2021; **49**(D1): D82–85.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Howe K, Chow W, Collins J, *et al.*: **Significantly Improving the Quality of Genome Assemblies through Curation**. *GigaScience*. 2021; **10**(1): g1aa153.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kumar S, Blaxter ML: **Simultaneous Genome Sequencing of Symbionts and Their Hosts**. *Symbiosis*. 2011; **55**(3): 119–26.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Lewin HA, Robinson GE, Kress WJ, *et al.*: **Earth BioGenome Project: Sequencing Life for the Future of Life**. *Proc Natl Acad Sci U S A*. 2018; **115**(17): 4325–33.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Oulhen N, Schulz BJ, Carrier TJ: **English Translation of Heinrich Anton de Bary's 1878 Speech, 'Die Erscheinung Der Symbiose' ('De La Symbiose')**. *Symbiosis*. 2016; **69**: 131–139.
[Publisher Full Text](#)
- Roeselers G, Newton ILG: **On the Evolutionary Ecology of Symbioses between Chemosynthetic Bacteria and Bivalves**. *Appl Microbiol Biotechnol*. 2012; **94**(1): 1–10.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Threlfall J, Blaxter M: **Launching the Tree of Life Gateway [version 1; peer review: not peer reviewed]**. *Wellcome Open Res*. 2021; **6**: 125.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Weis VM: **Cell Biology of Coral Symbiosis: Foundational Study Can Inform Solutions to the Coral Reef Crisis**. *Integr Comp Biol*. 2019; **59**(4): 845–55.
[PubMed Abstract](#) | [Publisher Full Text](#)